

## Perbandingan kinerja metode C4.5 dan Naive Bayes dalam klasifikasi artikel jurnal PGSD berdasarkan mata pelajaran

Utomo Pujiyanto<sup>1</sup>, Putri Yuni Ristanti<sup>2</sup>

1. Universitas Negeri Malang, Indonesia | utomo.pujiyanto.ft@um.ac.id
2. Universitas Negeri Malang, Indonesia | putriyuni7@gmail.com

### Abstrak

Pendidikan mempunyai standar sebagai acuan dalam proses pembelajaran. Dalam hal ini Pemerintah telah mengatur standar pendidikan di Indonesia, mengacu pada Peraturan Pemerintah Republik Indonesia Nomor 19 Tahun 2005 Pasal 6 ayat (1) yaitu kurikulum untuk jenis pendidikan umum, kejuruan, dan khusus pada jenjang pendidikan dasar dan menengah. Sesuai dengan Peraturan Pemerintah tersebut, ditetapkan Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 23 Tahun 2006 pasal 1 ayat (2), tentang Standar Kompetensi Lulusan yang diantaranya memuat SK-KMP (Standar Kompetensi Kelompok Mata Pelajaran). Standar inilah yang dijadikan sebuah rujukan untuk tenaga pendidik, dan bakal tenaga pendidik khususnya mahasiswa bidang pendidikan untuk membuat sebuah media pembelajaran, jurnal sebagai bahan ajaran yang pokok. Tujuan penelitian ini untuk mengklasifikasikan minat mahasiswa PGSD terhadap tema mata pelajaran menurut SK-KMP menggunakan metode Naive Bayes dan Decision tree J48. Hasil penelitian tersebut dapat dijadikan sebagai referensi untuk pengambilan tema pada mata pelajaran di tahun mendatang untuk lebih bervariasi, tidak hanya membahas tentang salah satu mata pelajaran tersebut. Kinerja dari kedua metode tersebut akan dibandingkan, sehingga dapat diketahui kinerja metode mana yang lebih baik dalam melakukan klasifikasi dokumen. Pengujian performa algoritma klasifikasi yang digunakan adalah teknik K-fold Cross Validation. Berdasarkan pengujian performa penerapan algoritma Naive Bayes dan Decision Tree J48 menggunakan teknik K-Fold Cross Validation terhadap 200 judul dan abstrak artikel jurnal, didapatkan algoritma Naive Bayes, tingkat akurasi sebesar 84%. Sementara itu, untuk hasil yang diperoleh dengan algoritma Decision Tree J48, tingkat akurasi sebesar 86%.

### Kata Kunci

Klasifikasi, SK-KMP, PGSD, Naive Bayes, Decision Tree J48, Weka

## 1. Pendahuluan

Kumpulan Jurnal yang terdapat didalam web Universitas kebanyakan hanya di urutkan berdasarkan volume dan tahun pembuatan jurnal. Dengan penataan konsep yang seperti ini, mahasiswa yang mencari tipe jurnal yang diinginkan, akan membuka satu persatu jurnal berdasarkan volume dan tahun, tentu hal ini sangat tidak efisien untuk mahasiswa. Penelitian ini memfokuskan kepada mahasiswa program studi PGSD untuk memudahkan pencarian jurnal berdasarkan SK-KMP untuk pengambilan tema pada mata pelajaran. Dengan Penelitian ini keuntungan yang didapatkan akan membantu mahasiswa prodi PGSD untuk dapat mendapatkan tipe jurnal berdasarkan mata pelajaran dengan sangat efisien dan pengambilan judul pada tahun depan untuk dapat lebih bervariasi, tidak hanya membahas tentang salah satu mata pelajaran. Konsep ini tidak merubah konsep penataan awal jurnal, tetapi melengkapinya dengan menambahkan fungsi klasifikasi. Ketika mahasiswa ingin mencari jurnal tetap merujuk pada volume dan tahun pembuatan, kumpulan jurnal yang ada didalam volume akan diklasifikasikan berdasarkan SK-KMP. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Naive Bayes* dan *Decision Tree* J48 pada klasifikasi jurnal PGSD.

Dengan berbagai informasi tersebut, informasi dapat diolah menggunakan sistem *Information Retrieval* dengan tujuan menemukan dokumen yang sesuai dengan kebutuhan seorang pengguna dengan lebih efisien. Dalam proses ini melibatkan tahap intermediate yaitu penyaringan, pencarian, operasi pencocokan data, dan peringkat (Huang et al., 2017). Beberapa penelitian dilakukan dengan menggunakan metode klasifikasi untuk mendapatkan sebuah informasi atau pengetahuan dari dataset tentang pendidikan, yaitu klasifikasi ke dalam kategori berdasarkan kualifikasi Pendidikan dengan menggunakan algoritma *naive bayes* memiliki akurasi 90% (Karthika & Sairam, 2015). Klasifikasi teks pada data siswa menggunakan algoritma *naive bayes* dengan akurasi 66,7 % dan KKN dengan akurasi 38.89 % (Rajeswari & Juliet, 2017), klasifikasi menggunakan *Decision Tree* J48 *Breast Cancer* dengan akurasi 95 % (Song & Woo, 2015). Penelitian ini hanya menggunakan satu data set yaitu jurnal PGSD, dalam penelitian ini akan dibandingkan kinerja dari metode *naive bayes* dan J48 dalam pengklasifikasian jurnal PGSD berdasarkan SK-KMP.

Dataset yang digunakan dalam penelitian ini berasal dari situs jurnal PGSD yang beralamat di <http://journal.student.uny.ac.id/jurnal/baca/99>. Atribut yang digunakan yaitu Volume dan tahun jurnal, Judul, dan Abstrak. Penelitian dilakukan dengan membandingkan algoritma *Naive Bayes* dan *Decision Tree* J48 untuk mengklasifikasi mata pelajaran berdasarkan SK-KMP. Hasil dari penelitian ini mahasiswa khususnya program studi PGSD dapat dijadikan sebagai referensi untuk pengambilan tema pada mata pelajaran di tahun mendatang untuk lebih bervariasi, tidak hanya membahas tentang salah satu mata pelajaran tersebut. Penelitian ini juga menghasilkan perbandingan kinerja dari masing-masing algoritma dengan melihat nilai *accuracy*, *precision*, dan *recall*

## 2. Metode

Pada penelitian ini menggunakan 2 algoritma yaitu *Naive Bayes* dan *Decision Tree* J48 untuk mengklasifikasikan jurnal PGSD, kinerja kedua algoritma tersebut diukur berdasarkan *accuracy*, *recall* dan *precision*.

### 1. Algoritma Naïve Bayes

Metode *Naive Bayes* merupakan metode yang populer dalam klasifikasi, metode ini terkenal dengan prediksi probabilitas dari data yang disajikan. *Naive Bayes* menggunakan teorema bayes. *Naive Bayes* memiliki nama lain seperti *idiot's Bayes*, *simple Bayes* dan *independence Bayes* karena kemudahannya serta tidak membutuhkan proses iterasi yang kompleks (Ashari, Paryudi, & Tjoa, 2013). *Naive Bayes* merupakan sebuah metode pengklasifikasian peluang sederhana dengan asumsi antar variabel penjelas saling bebas (*independen*) dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan (Suyadi, Setyanto, & Fattah, 2017). Keuntungan penggunaan Metode *Naive Bayes* hanya memerlukan data latih yang kecil untuk memperkirakan parameter *mean* dan *varians* dari variabel yang diperlukan untuk klasifikasi. Hal ini dikarenakan variabel independen yang diasumsikan hanya variabel untuk setiap kelas yang harus ditentukan, bukan seluruh matriks kovarians (Wahyuningtyas et al., 2015). *Naive Bayes* merupakan metode *supervised document classification* yang membutuhkan data *training* sebelum melakukan proses klasifikasi. Dalam prosesnya, *Naive Bayes* mengansumsikan ada tidaknya sebuah fitur pada suatu kelas tidak berhubungan dengan ada tidaknya fitur lain di kelas lain. Dalam proses pelatihan, dokumen telah ditentukan kategorinya. Kemudian akan diproses untuk membentuk pengetahuan berupa nilai probabilitas pada setiap *term* (kata). Proses ini akan menghasilkan sebuah *term* (kata) pada setiap dokumen yang mengkarakteristikan dokumen pada suatu kategori tertentu. Untuk menghitung probabilitas kategori dokumen dapat digunakan Pers. (1). sedangkan untuk menghitung setiap kata yang terdapat pada dokumen latih digunakan Pers. (2).

$$p(c_j) = \frac{n(doc_j)}{n(sampel)} \quad \text{Pers. (1)}$$

Keterangan:

$p(c_j)$  : probabilitas dokumen kategori  
 $n(doc_j)$  : jumlah seluruh dokumen pada suatu kategori  
 $n(sampel)$  : jumlah seluruh dokumen latih

$$p(w_i | c_j) = \frac{1 + n_i}{n + |x|} \quad \text{Pers. (2)}$$

Keterangan:

$p(w_i | c_j)$  : probabilitas kata pada setiap kategori

$n_i$  : frekuensi kemunculan kata pada setiap kategori

$n$  : jumlah seluruh kata dalam dokumen pada kategori tertentu

$|x|$  : jumlah total kata (*distinct*) di semua data latih

Tahap selanjutnya adalah mencari *term* (kata) dari dokumen yang belum diketahui kategorinya (data uji) dengan pengetahuan pada data latih  $P(w_i | c_j)$ . Kemudian hitung probabilitas setiap dokumen  $P(c_j)$  yang telah disimpan dipengetahuan pada saat proses pelatihan sebelumnya, maka untuk setiap kategori dokumen dapat dihitung menggunakan Pers. (3).

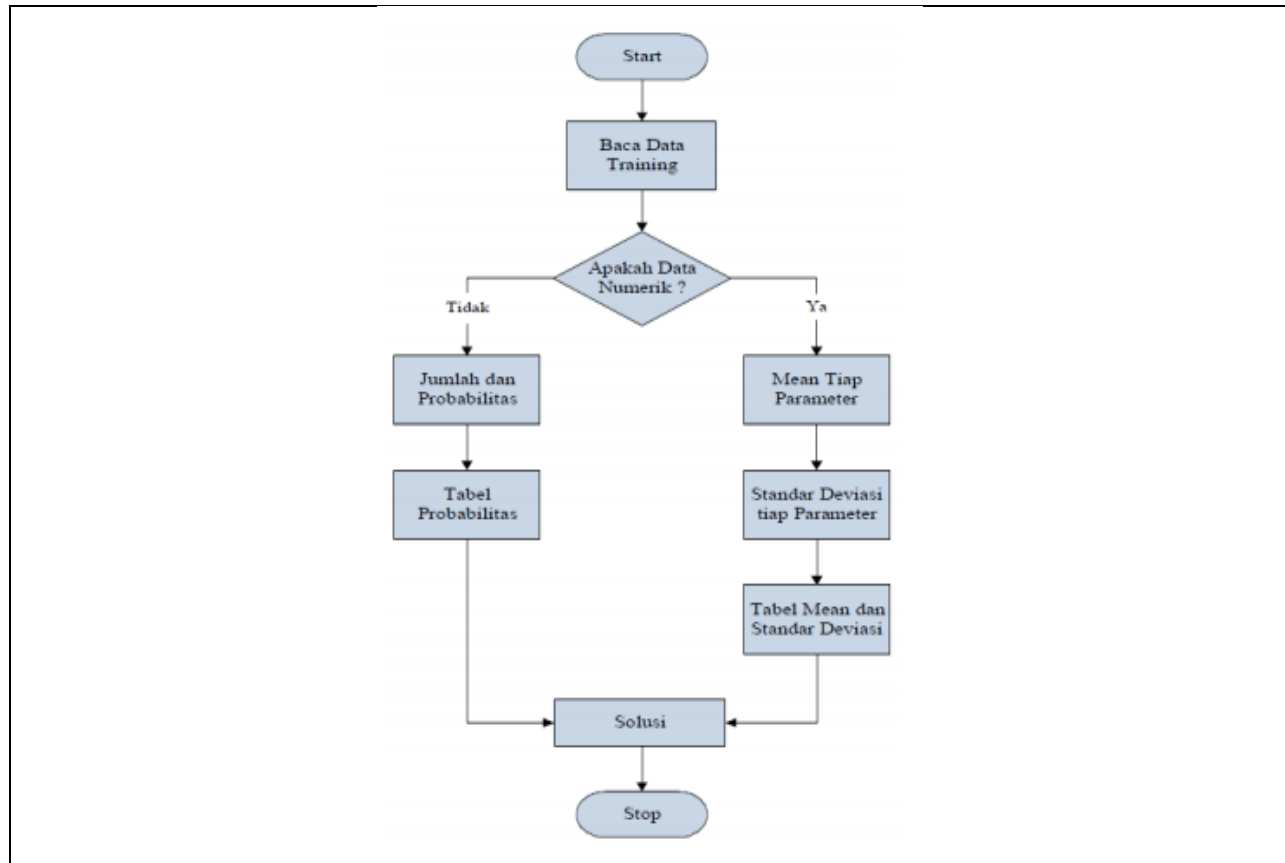
$$p(c_j) \prod_i p(w_i | c_j) \quad \text{Pers. (3)}$$

untuk mencari nilai tersebut, kalikan nilai  $p(w_i | c_j)$  yang merupakan nilai kumulatif perkalian probabilitas kemunculan kata yang sama pada data latih dengan nilai probabilitas dokumen yang sesuai kategorinya  $P(c_j)$ . Setelah didapatkan hasil perkalian pada masing-masing kategori dokumen, selanjutnya membandingkan dan mencari nilai probabilitas terbesar (*CMAP*) yang digunakan untuk klasifikasi data uji pada dokumen Jurnal PGSD yang akan diklasifikasikan ke dalam salah satu kategori yang tersedia (Schneider, 2005), perhitungan tersebut dapat dilihat pada Pers. (4).

$$c_{MAP} = \operatorname{argmax}_{c_j \in C} p(c_j) \prod_i p(w_i | c_j) \quad \text{Pers. (4)}$$

Alur dimulai dengan membaca data training, jika data numerik maka cari nilai *mean* dari setiap parameter, kemudian mencari standar deviasi. Sedangkan untuk mencari nilai dari probabilitas dengan menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data (Bustami, 2014).

Keunggulan lain dari Naïve Bayes antara lain memberikan hasil baik pada aplikasi yang melakukan operasi real-word data (Bilski, 2011), waktu yang dibutuhkan untuk komputasi relative singkat, kinerja klasifikasi lebih baik dengan menghilangkan fitur yang tidak sesuai, dan performanya bagus (Handayani, 2015). Sedangkan kelemahannya membutuhkan jumlah *record* yang sangat banyak untuk mendapatkan hasil yang baik serta kurang akurat jika digunakan pada dataset yang tidak sesuai (Jadhav & Channe, 2013). Algoritma Decision Tree J48



Gambar 1. Flowchart Naive Bayes

Algoritma berbasis *Decision Tree J48* merupakan implementasi *open source* Java dari algoritma C4.5. Sehingga J48 memiliki cara yang sama dalam membangun pohon dengan algoritma C4.5. Algoritma pembelajaran yang mengimplementasikan pencarian *heuristic* menggunakan pohon keputusan. Diawali dengan memilih atribut sebagai akar, selanjutnya membuat cabang dari setiap nilai. Kemudian membagi setiap kasus pada cabang pohon, dan mengulangi proses ini sampai keseluruhan cabang memiliki kasus yang sama. Algoritma ini didasarkan pada hierarki pohon yang digunakan untuk membangun pohon klasifikasi yang mewakili struktur sederhana dengan nodus non-terminal menunjukkan tes atribut dan nodus terminal untuk menghasilkan hasil keputusan. Untuk menghitung gain dalam memilih atribut yang menjadi root Pers. (5), sedangkan untuk menghitung nilai entropy digunakan Pers. (6).

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|s_i|}{|S|} * Entropy(s_i) \quad \text{Pers. (5)}$$

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

|S<sub>i</sub>|: jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

$$Entropy(S) = \sum_{i=1}^n p_i * \log_2 p_i \quad \text{Pers. (6)}$$

Keterangan:

S : himpunan kasus

A : fitur

n : jumlah partisi S

p<sub>i</sub> : proporsi dari S<sub>i</sub> terhadap S

## 2. Dataset yang digunakan

Dataset yang digunakan pada penelitian ini diambil dari [journal.student.uny.ac.id/jurnal/baca/99](http://journal.student.uny.ac.id/jurnal/baca/99) yang berisi data jurnal PGSD yang terdiri dari Volume dan tahun jurnal, Judul, dan Abstrak. Dataset merupakan data lama karena tahun diterbitkannya Jurnal PGSD tahun 2012 (Volume I) sampai tahun 2015 (Volume IV). Teknik pengambilan data dengan *Crawling* dari website resmi Universitas Negeri Yogyakarta.

Jumlah instance dari dataset jurnal PGSD berjumlah 200 data dengan 4 atribut ditampilkan pada Tabel 1.

**Tabel 1.** Daftar Atribut Dalam Dataset

Nama atribut	Penjelasan Atribut	Tipe data	Rentang nilai / enumerasi
Volume	Volume dan tahun jurnal tersebut di terbitkan secara resmi	String	kalimat
Judul	Judul dari jurnal PGSD UNY	String	kalimat
Abstrak	Berisi tentang penjabaran secara singkat tentang jurnal	String	kalimat
Mata Pelajaran	Atribut Kelas	nominal	Agama dan Akhlak Mulia, Kewarganegaraan dan Kepribadian, Ilmu pengetahuan dan Teknologi, Estetika, Jasmani, Olahraga dan Kesehatan

Untuk memberi label pada dataset yang diambil, menggunakan Dasar Kurikulum Peraturan Pemerintah Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan pasal 6 ayat (1) menyatakan bahwa kurikulum untuk jenis pendidikan umum, kejuruan, dan khusus pada jenjang pendidikan dasar dan menengah.

Metadata pemberian label dan penjelasan dari datasets yang telah diubah dengan teknik *string to words* ditampilkan pada tabel 2.

**Tabel 2.** Daftar Label Dalam Dataset

Label	Penjelasan Atribut	Tipe data	Rentang Nilai
Agama dan Akhlak	Cakupan yang merujuk pada jenis mata pelajaran	Numeric	0-100
Kewarganegaraan dan Kepribadian	Cakupan yang merujuk pada jenis mata pelajaran	Numeric	0-100
Ilmu Pengetahuan, dan Teknologi	Cakupan yang merujuk pada jenis mata pelajaran	Numeric	0-100
Estetika	Cakupan yang merujuk pada jenis mata pelajaran	Numeric	0-100
Jasmani Olahraga, dan Kesehatan	Cakupan yang merujuk pada jenis mata pelajaran	Numeric	0-100

Karakteristik dari dataset Jurnal PGSD yang digunakan dalam penelitian ini ditampilkan pada tabel 3.

**Tabel 3.** Karakteristik Dalam Dataset

Nama Kelas	Jumlah instance	Presentase
Kelas "Agama dan Akhlak"	6	3 %
Kelas "Kewarganegaraan dan Kepribadian"	22	11 %
Kelas "Ilmu Pengetahuan dan Teknologi"	168	84 %
Kelas "Estetika"	4	2 %

Pada Jurnal PGSD Universitas Negeri Yogyakarta tahun 2012 (Volume I) sampai tahun 2015 (Volume IV) tidak ada jurnal yang memuat cakupan yang merujuk mata pelajaran Jasmani, Olahraga dan Kesehatan. Sehingga peneliti tidak mencantumkan mata pelajaran tersebut dalam penelitian yang lebih lanjut.

### 3. Desain Eksperimen

WEKA adalah *workbench* untuk pembelajaran mesin dengan Bahasa pemrograman java. Aplikasi WEKA memungkinkan pengguna untuk mengidentifikasi informasi tersembunyi dari

database dan sistem file dengan pilihan dan antarmuka visual yang mudah digunakan. Workbench Weka berisi koleksi alat visualisasi dan algoritma misalnya (C4.5 (C5), ID3, K-means, dan Apriori) untuk memecahkan masalah data mining dan prediksi secara *real-world* (Kulkarni & Raj B. Kulkarni, 2016).

Klasifikasi adalah proses menemukan seperangkat model yang membahas dan membedakan ide dan kelas data, dengan tujuan untuk dapat menggunakan model tersebut untuk menebak kelas yang labelnya rahasia. Membangun model klasifikasi menggunakan data pelatihan. Model yang dihasilkan diuji dengan menugaskan label kelas ke kumpulan data objek data. Model diwakili sebagai pohon keputusan, aturan klasifikasi dan rumus matematika. Ini untuk mengklasifikasikan objek yang akan datang atau yang tidak diketahui. Tingkat ketepatan adalah persentase dari set pelatihan independen yaitu sampel uji yang diklasifikasikan dengan benar oleh model (Roshdi & Roohparvar, 2015).

Dalam penelitian ini peneliti menggunakan WEKA untuk melakukan pengklasifikasian jurnal PGSD. WEKA yang digunakan yaitu WEKA 3.8.2 dengan menggunakan algoritma yang disediakan, yaitu *Naive Bayes* dan *Decision Tree* J48. Spesifikasi perangkat keras yang digunakan yaitu processor Core i3 dengan memory 4 GB.

### 3.1 Langkah Penelitian

#### a) Data set

Proses awal penelitian dimulai dengan pengambilan data pada website resmi Universitas Negeri Yogyakarta dengan Crawling menggunakan Octoparse, kemudian file tersebut di export menjadi file excel.

#### b) *Pre processing*

Tujuan *pre processing* adalah mengubah data yang sebelumnya tidak terstruktur menjadi data yang terstruktur Tahap pada *pre processing* terbagi dalam beberapa tahap. Tahap pertama yaitu *case folding* yaitu Menghilangkan semua tanda baca pada teks data untuk menghilangkan anomali data sehingga data dapat diproses pada weka. Dataset yang sebelumnya diolah dalam excel .csv diubah menjadi bentuk file .arff. Pilih filter *StringToWordVector* untuk mengubah tipe data menjadi tipe numeric dan nominal. Edit parameter pada *StringToWordVector* yaitu parameter *lowerCaseTokens* menjadi *true* (mengubah per kata pada teks menjadi kecil) dan mengubah parameter *OutputWordCount* menjadi *true*. Tahap kedua yaitu Tokenization yaitu memotong kalimat (teks panjang) pada teks menjadi per kata. dengan mengubah parameter.

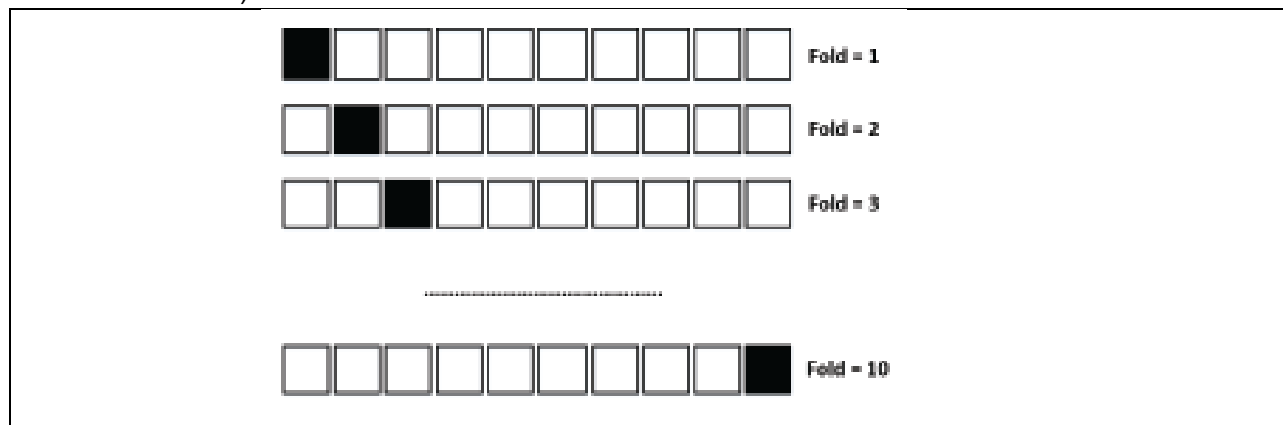
Selanjutnya proses *filtering* untuk menghilangkan *stopwords*. Tahap terakhir yaitu stemming, yaitu teks yang telah melalui proses *case folding* yang bertujuan untuk



menghilangkan imbuhan menjadi kata dasar (*root word*). Selanjutnya pilih *edit* untuk mengubah menjadi *set attribute as class*.

c) *Classification*

Proses *classification* dilakukan dengan mengimport dataset bentuk .arff dalam weka, disini peneliti menggunakan 2 algoritma yaitu *Naive Bayes* dan *Decisison Tree J48*. Pada pengklasifikasian ini peneliti menggunakan *Test Options Cross-Validations* yaitu salah satu metode validasi (biasa disebut *k-folds*). Kinerja dari *Cross-Validations*, dataset A diacak menjadi bagian subset  $A_1, A_2, A_3, A_4, \dots, A_k$  yang disebut *fold*. Setiap folds (kotak hitam pada **Gambar 2**) digunakan sebagai data uji, sedangkan fold lainnya digunakan sebagai data pelatihan (Lukito & Chrismanto, 2015).



**Gambar 2.** Ilustrasi *k-folds*

d) *Confusion Matrix*

Tahap ini dilakukan uji ketepatan dengan *confusion matrix* kedua algoritma, yaitu menghitung *accuracy*, *recall*, *precision* dan *error rate* dengan *Confusion Matrix* ditampilkan pada tabel 4.

**Tabel 4.** Tabel *Confusion Matrix*

Kelas	Terklarifikasi Positif	Terklarifikasi Negatif
Positif	TP ( <i>True Positive</i> )	FN ( <i>False Negative</i> )
Negatif	FP ( <i>False Positive</i> )	TN ( <i>True Negative</i> )

Pada kinerja *Confusion Matrix* dapat diukur dengan TP, FN, FP dan TN.

TP (*True Positive*) adalah jumlah data true yang yang memang memiliki nilai kebenaran true.

FN (*False Negative*) adalah jumlah data false yang memang memiliki nilai kebenaran false.

FP (False Negative) adalah jumlah data true yang dianggap sistem memiliki nilai kebenaran false.

TN (True Negative) adalah jumlah data false yang dianggap sistem memiliki nilai kebenaran true.

- a) Menghitung keakuratan sistem mengklasifikasikan data dengan tepat.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN}$$

- b) Menghitung seberapa banyak nilai kebenaran (positive) dari dataset yang memang bernilai benar (positive) muncul.

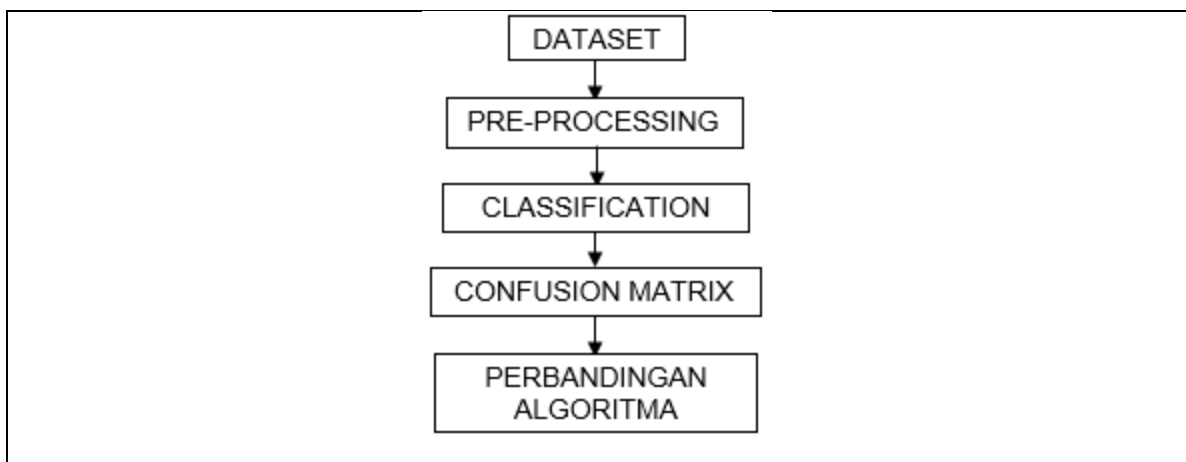
$$TPRate (Recall) = \frac{TP}{TP+FN}$$

- c) Ketepatan nilai kebenaran dalam klasifikasi sesuai dengan nilai kebenaran yang sesungguhnya.

$$Precision = \frac{TP}{TP+FP}$$

- e) Perbandingan Algoritma

Peneliti membandingkan kinerja algoritma *Naive Bayes* dan *Decisison Tree* J48 untuk mengetahui algoritma yang lebih baik untuk klasifikasi jurnal PGSD. Langkah-langkah penelitian dalam penelitian Information Retrieval ditunjukkan oleh gambar 3.



Gambar 3. Tahapan Penelitian

### 3. Hasil

Berikut adalah hasil dari penelitian membandingkan algoritma *Naive Bayes* dan *Decision Tree* J48 pada Jurnal PGSD Universitas Negeri Yogyakarta dengan *accuracy*, *presicion* dan *recall*.

#### a) Algoritma Naive Bayes

Hasil klasifikasi dataset jurnal PGSD dalam algoritma *Naive bBayes* dengan *test options* menggunakan *Cross-validation* dengan nilai *folds* 6 ditampilkan pada Tabel 5.

**Tabel 5.** Hasil Pengujian *Stratified Cross-validation*

<b>Stratified Cross-validation</b>		
<i>Summary</i>		
<i>Correctly Classified Instance</i>	168	84 %
<i>Incorrectly Classified Instance</i>	32	16 %
<i>Kappa statistic</i>	0.4019	
<i>Mean absolute error</i>	0.08	
<i>Root mean squared error</i>	0.2828	
<i>Relative absolute error</i>	54.7831 %	
<i>Root relative squared error</i>	106.6237 %	
<i>Total Number of Instance</i>	200	

Untuk hasil dengan *Confusion Matrix* pada algoritma *Naive Bayes* ditampilkan pada Tabel 6.

**Tabel 6.** Hasil *Confusion Matrix*

Agama dan akhlak mulia	Kewarganegaraan dan Kepribadian	Ilmu Pengetahuan dan teknologi	Estetika	<b>Classified as</b>
1	4	1	0	Agama dan akhlak mulia
2	10	10	0	Kewarganegaraan dan Kepribadian
0	10	157	1	Ilmu Pengetahuan dan teknologi
0	1	3	0	Estetika

Keterangan:

- Pada baris pertama nilai “1 4 1 0” menunjukkan bahwa instance class Agama dan akhlak mulia dalam data set JurnalPGSD, 5 salah diklasifikasikan sebagai Kewarganegaraan dan Kepribadian, dan 1 salah diklasifikasikan Ilmu Pengetahuan dan Teknologi.
- Pada baris kedua nilai “2 10 10 0” menunjukkan bahwa instance class Kewarganegaraan dan Kepribadian dalam data set JurnalPGSD, 2 salah diklasifikasikan sebagai Agama dan akhlak mulia, dan 10 salah diklasifikasikan Ilmu Pengetahuan dan Teknologi.
- Pada baris ketiga nilai “0 10 157 1” menunjukkan bahwa instance class Ilmu Pengetahuan dan Teknologi dalam data set JurnalPGSD, 10 salah diklasifikasikan sebagai Kewarganegaraan dan Kepribadian, dan 1 salah diklasifikasikan Estetika.

- d) Pada baris keempat nilai “0 1 3 0” menunjukkan bahwa instance class Estetika dalam data set JurnalPGSD, 1 salah diklasifikasikan sebagai Kewarganegaraan dan Kepribadian ,dan 3 salah diklasifikasikan sebagai Ilmu pengetahuan dan Teknologi.

Untuk hasil *Accuracy* pada algoritma Naive Bayes sebagai berikut:

$$accuracy = \frac{1+10+157+0}{200} 100\% = 84\%$$

Untuk hasil *Recall* pada algoritma Naive Bayes sebagai berikut:

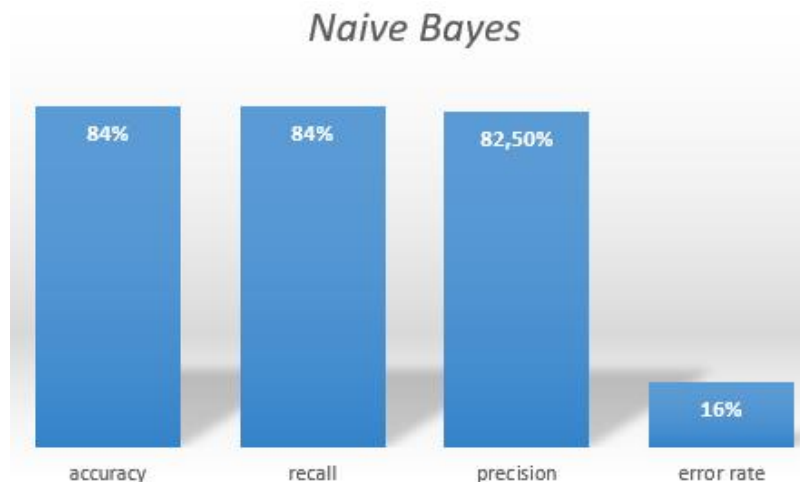
- a) Agama dan akhlak mulia =  $\frac{1}{1+(4+1+0)} = 0,167$   
 b) Kewarganegaraan dan Kepribadian =  $\frac{10}{10+(2+10+0)} = 0,455$   
 c) Ilmu Pengetahuan dan teknologi =  $\frac{157}{157+(0+10+1)} = 0,935$   
 d) Estetika =  $\frac{0}{0+(0+1+3)} = 0$

Nilai rata-rata *Recall* 84%

Untuk hasil *Precision* pada algoritma Naive Bayes sebagai berikut:

- a) Agama dan akhlak mulia =  $\frac{1}{1+(2+0+0)} = 0,333$   
 b) Kewarganegaraan dan Kepribadian =  $\frac{10}{10+(4+10+1)} = 0,400$   
 c) Ilmu Pengetahuan dan teknologi =  $\frac{157}{157+(1+10+3)} = 0,918$   
 d) Estetika =  $\frac{0}{0+(0+1+0)} = 0$

Nilai rata-rata *Precision* 82,5%



**Gambar 4.** Parameter Kinerja Algoritma *Naïve Bayes*

Hasil *Detailed Accuracy* Algoritma *Naive Bayes* pada Weka ditampilkan pada Tabel .

**Tabel 7.** *Detailed Accuracy Naive Bayes*

Parameter	Agama dan akhlak mulia	Kewarganegaraan dan Kepribadian	Ilmu Pengetahuan dan teknologi	Estetika	Weighted Avg
<i>TP Rate</i>	0,167	0,455	0,935	0,000	0,840
<i>FP Rate</i>	0,010	0,084	0,438	0,005	0,377
<i>Precision</i>	0,333	0,400	0,918	0,000	0,825
<i>Recal</i>	0,167	0,455	0,935	0,000	0,840
<i>F-Measure</i>	0,222	0,426	0,926	0,000	0,832
<i>MCC</i>	0,219	0,350	0,517	-0,010	0,480
<i>ROC Area</i>	0,724	0,782	0,776	0,480	0,769
<i>PRC Area</i>	0,221	0,309	0,925	0,020	0,818

## b) Algoritma *Decisison Tree* J48

Hasil klasifikasi dataset jurnal PGSD dalam algoritma *naive bayes* dengan *test options* menggunakan *Cross-validation* dengan nilai *folds* 6 ditampilkan pada Tabel 8.

**Tabel 8.** Hasil Pengujian *Stratified Cross-validation*

<b>Stratified Cross-validation</b>		
<i>Summary</i>		
<i>Correctly Classified Instance</i>	174	87%
<i>Incorrectly Classified Instance</i>	26	13 %
<i>Kappa statistic</i>	0.4488	
<i>Mean absolute error</i>	0.0831	
<i>Root mean squared error</i>	0.2469	
<i>Relative absolute error</i>	56.8742 %	
<i>Root relative squared error</i>	93.065 %	
<i>Total Number of Instance</i>	200	

Untuk hasil dengan *Confusion Matrix* pada algoritma *Naive Bayes* ditampilkan pada Tabel 9.

**Tabel 9.** Hasil *Confusion Matrix*

Agama dan akhlak mulia	Kewarganegaraan dan Kepribadian	Ilmu Pengetahuan dan teknologi	Estetika	<b>Classified as</b>
4	0	2	0	Agama dan akhlak mulia
2	7	13	0	Kewarganegaraan dan Kepribadian
0	4	163	1	Ilmu Pengetahuan dan teknologi
0	1	2	0	Estetika

Keterangan:

- a) Pada baris pertama nilai “4 0 2 0” menunjukkan bahwa instance class Agama dan akhlak mulia dalam data set JurnalPGSD, 2 salah diklasifikasikan sebagai Ilmu Pengetahuan dan teknologi
- b) Pada baris kedua nilai “2 7 13 0” menunjukkan bahwa instance class Kewarganegaraan dan Kepribadian dalam data set JurnalPGSD, 2 salah diklasifikasikan sebagai Agama dan akhlak mulia, 13 salah diklasifikasikan Ilmu Pengetahuan dan Teknologi.
- c) Pada baris ketiga nilai “0 4 163 1” menunjukkan bahwa instance class Ilmu Pengetahuan dan Teknologi dalam data set JurnalPGSD, 4 salah diklasifikasikan sebagai Kewarganegaraan dan Kepribadian, dan 1 salah diklasifikasikan Estetika.
- d) Pada baris keempat nilai “0 2 2 0” menunjukkan bahwa instance class Estetika dalam data set JurnalPGSD, 2 salah diklasifikasikan sebagai Kewarganegaraan dan Kepribadian, dan 2 salah diklasifikasikan sebagai Ilmu Pengetahuan dan teknologi.

Untuk hasil *Accuracy* pada algoritma *Decision Tree* J48 sebagai berikut:

$$accuracy = \frac{4+7+163+0}{200} 100\% = 86 \%$$

Untuk hasil *Recall* pada algoritma *Decision Tree* J48 sebagai berikut:

- a) Agama dan akhlak mulia =  $\frac{4}{4+(0+2+0)} = 0,667$
- b) Kewarganegaraan dan Kepribadian =  $\frac{7}{7+(2+13+0)} = 0,318$
- c) Ilmu Pengetahuan dan teknologi =  $\frac{163}{163+(0+4+1)} = 0,970$
- d) Estetika =  $\frac{0}{0+(0+1+2)} = 0$

Nilai rata-rata *Recall* 87%

Untuk hasil *Presicion* pada algoritma *Decision Tree* J48 sebagai berikut:

- a) Agama dan akhlak mulia =  $\frac{4}{4+(2+0+0)} = 0,667$
- b) Kewarganegaraan dan Kepribadian =  $\frac{7}{7+(0+4+1)} = 0,538$
- c) Ilmu Pengetahuan dan teknologi =  $\frac{163}{163+(2+13+2)} = 0.906$
- d) Estetika =  $\frac{0}{0+(0+0+1)} = 0$

Nilai rata-rata *Precision* 84%



**Gambar 5.** Parameter Kinerja Algoritma *Decision Tree J48*

Hasil *Detailed Accuracy* Algoritma *Naive Bayes* pada Weka ditampilkan pada Tabel 10.

**Tabel 10.** Hasil *Confusion Matrix*

Parameter	Agama dan akhlak mulia	Kewarganegaraan dan Kepribadian	Ilmu Pengetahuan dan teknologi	Estetika	Weighted Avg
<i>TP Rate</i>	0,667	0,318	0,970	0,000	0,870
<i>FP Rate</i>	0,010	0,034	0,531	0,005	0,450
<i>Precision</i>	0,667	0,538	0,906	0,000	0,840
<i>Recall</i>	0,667	0,318	0,970	0,000	0,870
<i>F-Measure</i>	0,667	0,400	0,937	0,000	0,851
<i>MCC</i>	0,656	0,361	0,536	-0,010	0,510
<i>ROC Area</i>	0,747	0,661	0,745	0,480	0,724
<i>PRC Area</i>	0,622	0,300	0,910	0,020	0,817

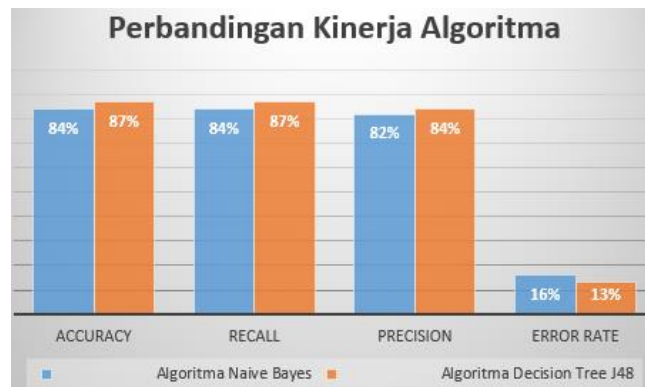
### c) Perbandingan Algoritma *Naïve Bayes* dan *Decision Tree J48*

Setelah melakukan penelitian dalam mengklasifikasikan dataset dengan menggunakan algoritma *naïve bayes* dan *Decision Tree J48*, hasil pengujian berdasarkan *accuracy*, *recall*, *persicion*, dan *error rate* ditampilkan pada Tabel 11.

**Tabel 11.** Hasil *Confusion Matrix*

Perbandingan	Algoritma	
	Naive Bayes	Decision Tree J48
<i>accuracy</i>	84%	87%
<i>recall</i>	84%	87%
<i>precision</i>	82%	84%
<i>Error Rate</i>	16%	13%

Berikut grafik perbandingan kinerja kedua algoritma berdasarkan *accuracy*, *recall*, *precision*, dan *error rate* ditampilkan pada Gambar 9



Gambar 9. Parameter Kinerja Kedua Algoritma

#### 4. Kesimpulan

Berdasarkan uraian hasil penelitian, maka dapat dikemukakan kesimpulan sebagai berikut:

- Berdasarkan hasil penelitian klasifikasi Jurnal PGSD dengan menggunakan algoritma *naive bayes* dan algoritma *Decision Tree J48*, hasil *accuracy* dari kedua algoritma tersebut menunjukkan tingkat *accuracy Decision Tree J48* sebesar 87 % lebih baik dari pada algoritma *Naive Bayes* sebesar 84%.
- Algoritma *Decision Tree J48* juga memberikan nilai pada *recall* dan *precision* sebesar 87% dan 84 % lebih baik dibandingkan algoritma *Naive Bayes* dengan nilai 82% dan 83%.
- Dengan hasil parameter pengujian menggunakan confusion matrik, peneliti merekomendasikan penggunaan algoritma *Decision Tree J48* pada klasifikasi Jurnal PGSD. Hasil *accuracy* akan menunjukkan hasil berbeda tergantung dengan jenis data, nilai *folds* dan jumlah *instance*, label *class*.

#### Daftar Rujukan

- Ashari, A., Paryudi, I., & Tjoa, A. (2013). Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool. *International Journal of Advanced Computer Science and Applications*, 4(11), 33–39. <https://doi.org/10.14569/IJACSA.2013.041105>
- Bilski, A. (2011). A review of artificial intelligence algorithms in document classification.



*International Journal of Electronics and Telecommunications*, 57(3), 263–270.  
<https://doi.org/10.2478/v10177-011-0035-6>

Bustami. (2014). Penerapan Algoritma Naive Bayes. *Jurnal Informatika*, 8(1), 884–898.

Handayani, F. dan F. S. P. (2015). Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110, 7(1).

Huang, F., Zhu, Q., Zhou, J., Tao, J., Zhou, X., Jin, D., ... Wang, L. (2017). Research on the parallelization of the DBSCAN clustering algorithm for spatial data mining based on the Spark platform. *Remote Sensing*, 9(12). <https://doi.org/10.3390/rs9121301>

Iqbal, M., Abid, M. M., Waheed, U., & Kazmi, S. H. A. (2017). Classification of Malicious Web Pages through a J48 Decision Tree, a Naïve Bayes, a RBF Network and a Random Forest Classifier for WebSpam Detection. *International Journal of U- and E- Service, Science and Technology*, 10(4), 51–72. <https://doi.org/10.14257/ijunesst.2017.10.4.05>

Jadhav, S. D., & Channe, H. P. (2013). Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques. *International Journal of Science and Research (IJSR)*, 14611(1), 2319–7064. Retrieved from [www.ijsr.net](http://www.ijsr.net)

Karthika, S., & Sairam, N. (2015). A Naïve Bayesian Classifier for Educational Qualification. *Indian Journal of Science and Technology*, 8(16).  
<https://doi.org/10.17485/ijst/2015/v8i16/62055>

Kulkarni, E. G., & Raj B. Kulkarni. (2016). WEKA Powerful Tool in Data Mining. *International Journal of Computer Applications*, 975 – 888(8887), 10–15.

Lukito, Y., & Chrismanto, A. R. (2015). Perbandingan Metode-Metode Klasifikasi Perbandingan Metode - Metode Klasifikasi Untuk Indoor Positioning System. *Jurnal Teknik Informatika Dan Sistem Informasi (JuTISI)*, 1(AUGUST 2015), 123–131. Retrieved from <http://jutisi.maranatha.edu/index.php/jutisi/article/download/373/368>

Rajeswari, R. P., & Juliet, K. (2017). Text Classification for Student Data Set using Naive Bayes Classifier and KNN Classifier. *International Journal of Computer Trends and Technology*, 43(1), 8–12.

Roshdi, A., & Roohparvar, A. (2015). Review : Information Retrieval Techniques and Applications. *International Journal of Computer Networks and Communications Security*, 3(9), 373–377.

Schneider, K.-M. (2005). Techniques for Improving the Performance of Naive Bayes for Text Classification. *Computational Linguistics and Intelligent Text Processing*, (i), 682–693.  
[https://doi.org/10.1007/978-3-540-30586-6\\_76](https://doi.org/10.1007/978-3-540-30586-6_76)

- Song, J. H., & Woo, H. Y. (2015). A study on AQ (adversity quotient), job satisfaction and turnover intention according to work units of clinical nursing staffs in Korea. *Indian Journal of Science and Technology*, 8(November), 74–78. <https://doi.org/10.17485/ijst/2015/v8i>
- Suyadi, S., Setyanto, A., & Fattah, H. Al. (2017). Analisis Perbandingan Algoritma Decision Tree (C4.5) Dan K-Naive Bayes Untuk Mengklasifikasi Penerimaan Mahasiswa Baru Tingkat Universitas. *Indonesian Journal of Applied Informatics*, 2(1), 59. <https://doi.org/10.20961/ijai.v2i1.13258>
- Wahyuningtyas, R. S., Pratiwi, H. S., Studi, P., Informatika, T., Teknik, F., & Tanjungpura, U. (2015). Sistem Pakar Penentuan Jenis Kulit Wajah Wanita Menggunakan Metode Naïve Bayes, 1(1).